DOCUMENT RESUME

ED 222 498 TM 820 390

AUTHOR Rice, G. Elizabeth; Higgins, Norman
TITLE The Role of Standardized Testing in the Teacher's Evaluation Strategy.
SPONS AGENCY Arizona State Univ., Tempe. Coll. of Education.
PUB DATE Mar 82
NOTE 29p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New York, NY, March 20-22, 1982).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Classroom Observation Techniques; *Classroom Techniques; Elementary Education; Ethnography; Evaluation Methods; Examiners; Program Effectiveness; *Standardized Tests; Student Evaluation; *Teacher Attitudes; Teacher Effectiveness; *Testing
IDENTIFIERS *California Achievement Tests

ABSTRACT

Teachers' perspectives on the student assessment process and on the role of testing in that process are presented. Primarily ethnographic in approach, this study includes investigations of teachers' ideas and thought processes as well as observation of actual classroom practice. Six elementary teachers participated in a three-stage process of initial interviews, classroom observation, and post-interviews. Interviews included standard ethnosemantic procedures for eliciting a taxonomy of evaluation techniques, questions about the uses to which these techniques are put, and a judged similarity task for eight commonly mentioned techniques. The classroom observation sessions included the administration of the California Achievement Tests. Results indicate that there is considerable variation in teachers' opinions about standardized testing and in the classroom procedures for administering these tests. Despite this variation, the actual degree to which this testing and its products are employed by teachers is consistently quite small. Teachers perceive as most valid those techniques which are fully integrated into classroom instruction and over which they have full control. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED222498

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

The Role of Standardized Testing in the Teacher's Evaluation Strategy

G. Elizabeth Rice

Norman Higgins

College of Education
Arizona State University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
G. E. Rice
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper prepared for a symposium, The Practice of Standardized Testing, NCME, New York, March, 1982. Research supported by the Schooling Practices Laboratory, College of Education, Arizona State University. The authors extend their thanks to the other members of the research team - Camille Casteel, Carole Edelsky, Maurine Fry, Sharen Kellogg, David Lancy, Wilhelmina Savenye, Ann Schmidt, Julie Stout, and the six teacher participants.

TM 820 390

ABSTRACT

Standardized testing is one of an array of evaluation techniques available to a teacher. The goal of the present study is to describe the teacher's perspective on the process of evaluating students and on the role of testing in that process. The study is primarily ethnographic in approach, and includes investigations of teacher's ideas and thought processes as well as observation of actual classroom practice. Six elementary teachers participated in a three-stage process of initial interviews, classroom observation, and post-interviews. Interviews included standard ethnosemantic procedures for eliciting a taxonomy of evaluation techniques, questions about the uses to which these techniques are put, and a judged similarity task for eight commonly mentioned techniques. The classroom observation sessions included the administration of the California Achievement Test. Results indicate that there is considerable variation in teachers' opinions about standardized testing and in the classroom procedures for administering these tests. Despite this variation, the actual degree to which this testing and its products are employed by teachers is consistently quite small. Teachers perceive as most valid those techniques which are fully integrated into classroom instruction and over which they have full control.

INTRODUCTION

The increased use of testing in the schools for educational accountability -- determining student competence and teacher/program effectiveness -- has stimulated considerable research regarding teachers' perceptions and uses of tests in the classroom. Much of this research has focused on the classroom teacher's use of standardized achievement tests such as the California Test of Basic Skills (CTBS) and the California Achievement Test (CAT). Three major findings that have emerged from this research: 1) there is often a mismatch between what is taught in the classroom and what is assessed on standardized tests (Hanson & Schutz, 1980; Leinhardt & Seewald, 1980; Cole & Nitko, 1979; Rosenshine, 1978); 2) teachers report that standardized achievement tests are of very little use to them beyond confirming their perceptions of student abilities (Salmon-Cox, 1980); and 3) teachers use a variety of nonstandardized student assessment techniques in their day-to-day teaching activities (Salmon-Cox, 1980; Yeh, 1978).

In this paper we will describe how the standardized achievement test fits into the teacher's overall strategy for evaluating student performance. Data from interviews with teachers and from observation of classroom testing procedures are presented. These include teachers' comparisons of achieve-

ment tests with other evaluation techniques as well as teachers' opinions about and uses for standardized tests. An analysis of variation in procedures for administering standardized tests is also presented.

The research on which the paper is based is part of a more general study of teachers' perceptions of the process of evaluating student performance and of the role of testing in that process. That descriptive study was undertaken to provide baseline data for an evaluation of Arizona's state-mandated testing programs as they are being implemented in a local school district. Standardized achievement tests were not the main focus of the study, which had as its specific goals 1) to develop a taxonomy of techniques that teachers use to assess students, 2) to describe teachers' uses and perceptions of various assessment techniques, and 3) to describe the implementation of various assessment activities in natural classroom settings.

METHODS

The methods used in this study were designed to provide a description of the teacher's perspective on the process of assessing students and on the role of testing (including standardized achievement testing) in that process. This includes investigations of teachers' ideas and thoughts about assessment

techniques as well as observation of actual classroom practice. The methods used are primarily ethnographic in derivation. Such methods are particularly appropriate in a descriptive study because they have been developed to enable the observer researcher to describe situations as perceived by the people being studied with as little observer bias as possible. In this study, a three stage design of initial interviews, classroom observations, and post-interviews was used.

Participants

Participants in the study included six teachers from the Chandler Schools' Unified District and their fourth, fifth and sixth grade classes. Teachers were chosen to participate in the study based on nominations from their principals as "good" teachers. Participation was completely voluntary. The teachers were paid fifty dollars for their participation. In addition to the teacher-participants, four observer/interviewers were trained to collect data. Two of these were graduate students at Arizona State University, and two were classroom teachers on temporary assignment in the Chandler Schools' District Office.

Procedures

Each teacher was matched with an interviewer/observer, who conducted all the interviews and observations with that teacher. While the interviewer/observers were not able to spend suffi-

cient time in the field to produce an ethnographic account, this procedure did allow for the development of rapport between teacher and researcher.

Interviews. Two interviews were conducted with each teacher. The initial interview was open-ended and ethnographic in format (Spradley, 1979) and was designed to elicit a taxonomy of evaluation techniques used by the teachers as well as situations in which these techniques are employed. The teachers were asked to describe the methods or techniques which they use to assess the performance of their students, and to describe in detail the classroom procedures followed for each of the techniques. They were also asked to list the situations in which they would want or would be asked to assess the progress of their students. They were then asked to match these situations with the techniques they had described. The questions were deliberately non-directive in nature so as to allow the teacher's own view of the assessment process to be expressed. This initial interview took between one and three hours to conduct.

The second interview was conducted after the completion of the classroom observation sessions (described below). This interview was more structured in format. The teachers were asked to rate the validity of eight of the assessment techniques most commonly mentioned in the initial interviews. They were asked to compare each assessment technique in a triads compari-

son task, identifying which technique was different from the other two techniques presented in the triad. In this interview, teachers also rated their agreement with various statements of opinion taken from the original interviews. This interview lasted about one hour.

Classroom Observations. Between the two interviews, the interviewer/observer assigned to each teacher conducted about ten hours of classroom observation, generally in five two-hour sessions. The major purpose off the observation sessions was to collect descriptions of actual classroom examples of the evaluation techniques which the teachers had described in the initial interview. Since the timing of the observation sessions coincided with the state-wide administration of the California Achievement Tests, the administration of at least one section of this standardized test by each teacher was observed. Observers also took notes on the physical and social characteristics of the classrooms. For purposes of observer training, data control, and establishing rapport with the teachers, the project anthropologist (Rice) also accompanied the interviewer/observer to one observation session with each classroom.

Analysis

In addition to the general compilation of responses from the interviews, two specific analysis procedures were used: taxonomic analysis and multidimensional scaling.

Taxonomic Analysis. Based on responses to the intitial interview, a taxonomy of evaluation techniques was prepared for each teacher. A taxonomy is an outline which shows the hierarchical or inclusion relations among a set of items (see Spradley, 1979). These taxonomies were presented to the teachers at the final interview for verification and/or modification. The individual taxonomies were then combined into a summary taxonomy of assessment techniques which appears as Figure 1.

Multidimensional Scaling. The responses to the triads comparison task of eight commonly mentioned assessment techniques were compiled into a similarity matrix. This matrix provides a numerical representation of how similar or different the teachers perceive each pair of the eight assessment techniques to be. The matrix was submitted to a multidimensional scaling program (ALSCAL) which provides a visual representation of these similarity measurements. A two dimensional solution is presented in Figure 2.

RESULTS AND INTERPRETATION

Results from the study will be presented in two sections. In the first section we will attempt to show how standardized achievement tests fit into teachers' overall strategies for

evaluating students. This will involve comparison of achievement tests to other evaluation techniques and a brief exploration of teachers' opinions about and uses for these tests. In the second section we will consider results from our classroom observations. The second section presents an analysis of sources of variation in classroom procedures for administering the test.

Achievement Testing and Other Evaluation Strategies

The data from this study indicate that teachers make use of a wide variety of techniques when evaluating the performance of their students. The taxonomy in Figure 1 presents an outline of the techniques mentioned by the six teachers interviewed. The items on the left of the taxonomy are the more general categories; those on the right represent specific evaluation techniques. The teachers mentioned an average of fifteen techniques during our interviews with them. As can be seen from the taxonomy, three broad categories of evaluation emerged from the interviews: observation, assigned work, and tests. Within these categories, the one most elaborated (that is, with the most specific techniques mentioned) is assigned work, and specifically classwork. In general, the most elaborated categories in a taxonomy are the ones which are most salient or important to the responder. The significant role of classwork in the life

of the teacher is confirmed by our classroom observations, which indicate that the bulk of a teacher's "typical" day is spent conducting classwork.

It is clear from this taxonomy that the various tests available to teachers constitute only a part of their overall student assessment information. Similarly, standardized achievement tests are only one type of test available to teachers. During the interviews, only half of the teachers mentioned the achievement tests, and one of these denied that it was a technique for evaluating students at all. While achievement tests do appear as part of teachers' repertoire for evaluating students, it became clear in the interviews that such tests play a relatively small role in that process.

Another procedure which we used to explore the way achievement tests fit into the teacher's repertoire of evaluation techniques was a multimensional scaling of eight commonly mentioned evaluation techniques. As described in the Methods section, the multidimensional scaling is based on teachers' reponses to a triads task. The analysis, which appears in Figure 2, gives a graphic representation of teachers' perceptions of how similar these evaluation techniques are. The closer together two techniques are on the graph, the more similar they are perceived to be. Conversely, techniques which are far apart on the graph are perceived to be different from one another. In this analysis, for example, observation and dis-

cussion are perceived by the teachers in our sample as being very similar to each other, while observation and the California Achievement Test are very different. The two dimensional solution accounts for 98 percent of the variance in the responses.

The interpretation of the dimensions on this graph comes from the teachers themselves. Teachers were asked to explain each of their choices on the triads comparison tasks, and the features which they mentioned were compiled. Based on these features, the long, horizontal axis of the matrix represents a "formality" dimension. The techniques on the rights side are the least formal; those on the left are most formal. Thus, classwork, observation, and discussion can be labelled as relatively informal techniques, while program tests, district competency tests, and the California Achievement Test (CAT) are formal evaluation techniques. Teacher-made tests occupy a midposition, tending toward the formal end. Similarly, the vertical axis represents a dimension of "curriculum relevance." Along this dimension, teacher-made tests, program tests and classwork are most relevant to the instructional program which the teacher is presenting. Observation and discussion seem to be perceived as somewhat less related to the curriculum, probably because they are more fluid, can go at tangents to the actual program of instruction, and may serve purposes other than evaluation or instruction. The standardized achievement test, here represented by the CAT, is perceived as being least related

to classroom instruction.

A concern with the relevance of the achievement test for the curriculum was a persistent theme among teachers. When attempting to differentiate the CAT from the other evaluation techniques five of the six teachers mentioned this problem. Some typical comments are these: (the CAT) is "not closely related to classroom program"; is "imposed from outside--not teacher's view of what students should know"; "gives less information than teacher controlled methods."

The multidimensional scaling technique requires that teachers make comparative judgements about the evaluation techniques, but such judgements are not necessarily value judgements. To determine how valuable teachers felt these eight techniques are, we asked them to provide rankings. Teachers were asked to rank the techniques from 1 (best) to 8 (worst) as to which "provides the best evaluation of what a student knows." The rankings for each teacher are presented in Table 1. No measures of central tendency are provided for these rankings, since there is a clear bimodal distribution. Teachers A, B, and C rank teacher-made tests as the best assessment technique, while teachers D, E, and F prefer observation. Discussion, classwork and program tests appear to constitute the next class of techniques, with the competency tests and the standardized test (CAT) being judged as the least effective technique. Note that only one teacher (Teacher A) judged the CAT to rank in the

top half of the list.

The even split which is apparent in the rankings between those teachers who prefer observation and those who prefer semiformal assessment techniques (e.g. teacher-made tests), is confirmed by other data. Responses to the three statements of opinion which appear in Table 2 are an example. When asked to judge the statement "The CAT (California Achievement Test) is mass cruelty" (this is an actual quote taken from an interview with one of the teachers), three teachers agreed; three disagreed. A similar and complementary pattern appeared for the statement "there is no better way to assess students than by talking to them" (another quote from a teacher): half of the teachers agreed strongly, the other half disagreed. Those who favor informal techniques are not necessarily opposed to testing, however, no teachers disagreed with the statement "tests provide an objective measurement of students' knowledge," although two were neutral to this statement. This division between those teachers who believe the CAT to be "cruel" and those who do not is also related to differences in procedure in administering the test, as we shall see below.

With respect to teachers' use of information from standardized tests, only three teachers reported any uses of this information during the interviews. Keep in mind, however, that we were asking teachers only about situations in which they were asked to evaluate students, so that other possible uses of the

test were not investigated. With respect to evaluation uses, two teachers reported using results of previous years' tests for grouping of students, one teacher reported discussing results in parent conferences. Two teachers said that they used CAT results to "get a general picture of the students and how they have done," "but not for grading or anything." One teacher reported using results for self-evaluation, and two teachers believed that the School District used the tests to evaluate them, rather than the students. The majority of teachers did not use CAT results to evaluate their students, and those who did use them to some degree report that they "never take achievement test results as the first or final word on how a student is doing."

As indicated earlier, teachers' misgivings about standardized achievement tests stem largely from discrepancies between the tests and the instructional program. Teachers commented that test instructions are too difficult, are ambiguous, or do not correspond to the way in which information has been presented in the classroom. Sometimes the tests will require skills which have not yet been covered in the classroom. The special formality of the testing situation is also grounds for considering results to be of limited value, in some teachers' view: students may be uncomfortable and anxious with the unfamiliar procedures. They may also just have a "bad day." One teacher suggested that the CAT was good for evaluating how

well a student was doing "in test taking."

To summarize the results of this section, the standardized achievement test is not a particularly salient evaluation technique for teachers. They do not believe that such tests are particularly good for evaluating what a student knows, and they perceive them to be highly formal and not clearly related to the instructional program.

Classroom Procedures for Administering Achievement Tests

We observed six teachers administering various portions of the California Achievement Test. The State of Arizona required that this test be given to all students during the third week in April, but teachers were free to choose their administration times during that period. All six of our teachers chose to spread the test over several days. We observed at least half an hour of testing in each class.

There was a great deal of similarity among teachers in some aspects of their administration procedures, and wide variation in other aspects. Our analysis of teachers' classroom procedures in administering the CAT indicates that they follow a sequence of eight major activities. These activities are conducted in the following order: arrange seating, give preinstructions, distribute materials, give test directions, begin, proctor, stop, and give a break. Except for the preinstructions

and the break, these activities are conducted by all the teachers in the same order. Given the constraints of the formalized testing situation, deviations from this general order of procedure would be very unlikely, particularly with an observer present.

The sequence of activities is depicted in Figure 3. Specific actions taken by the teachers in each of these categories are listed under each of the eight major activities. Those actions which were performed by all teachers have been underlined in the figure. The other actions are followed by a ratio which gives the number of teachers observed performing this action over the number of cases in which field notes were complete enough to determine if the action occurred. Thus, all teachers arranged the classroom seating so that desks were separated from each other (in half the classes this required rearranging the normal seating pattern), had students clear their desks, distributed necessary testing materials, read the printed directions, including the sample problem, told students to begin, figured the stop time, proctored so as to suppress distraction and prevent cheating, stopped the students, and collected the test materials. This is the basic and common administration procedure.

Variation in testing procedures came primarily in the instructional and proctoring activities of the teachers. Three teachers gave "preinstructions" before passing out materials.

These included specific test taking strategies, such as "if you don't know the answer, guess" or "if you can't work a problem, go on to the next," as well as classroom rules for testing situations, such as "no talking" or "no passes to the bathroom." The two teachers giving strategies gave opposite advice on whether or not to guess at the answers to problems on the test. Later discussions with teachers indicated that some of them believe that teaching test taking strategies to students makes a significant improvement in their performance and that they begin doing so a week or so before the tests are given, so that our observed proportion of teachers suggesting strategies for students is probably lower than the actual one. This suggests that a complete description of achievement test administration would require observers in the classroom well in advance of testing dates.

The instructional segment of the testing procedure always includes the teacher reading the printed directions from the test and giving the answers to the sample questions. The amount of question answering and explanation of directions provided beyond this varies from none to several minutes of explanation. Two teachers spent time defining words such as "opposite" and "run-on" from the test directions for their students. Three teachers moved around the room to see if the students had done the sample problem correctly, and three encouraged their students to "do their best." During this instructional segment,

three teachers commented on the difficulty of the test, and only one on the need to work quickly. Students are thus exposed to a variety of explicit and implicit messages from the teacher about the content of the test and the meaning of the testing situation.

Proctoring activities are another source of variation in testing procedures. While all teachers made efforts to suppress distractions by reducing noise inside and outside the classroom to a minimum, and to prevent cheating by watching the class and prohibiting students from talking and looking at other's tests, other proctoring activities varied widely. The amount of time the teacher spent moving around the room varied from none - the teacher remained stationary - to constant movement around student desks. This quantity of motion was directly related to the amount of help the teacher gave to students. The number of students with hands raised visited by teachers in a half hour period ranged from zero to 25, and the number of students at whose desks teachers stopped without being summoned by the student ranged from one to twelve. These exchanges with students were conducted in hushed voices and commonly included such comments as "I can't tell you the answer," "do the best you can," "go on to the next one." No instances of substantive help were heard by the observers, though one teacher, discovering a common misunderstanding, stopped the class to explain the directions within the test itself. Other actions in which

single teachers engaged include announcing the time left four minutes from the end of the test, and passing out papers and worksheets unrelated to the test while students were working.

The amount of teacher activity during the proctoring segment of the testing appears to be related to the teacher's opinions about the test. As was mentioned above, three teachers agreed with the statement "the CAT is mass cruelty." All of these teachers made themselves available to the students for help and encouragement during the testing. They were in constant motion around the classroom and visited students at a rate about twice that of the other teachers. While this may be seen as an attempt to mitigate the "cruelty" of the testing situation, it also reflects the more interactive style of these classrooms. That is, students have come to expect a certain degree of interaction and help from the teacher, and these expectations are reflected in the number of hands raised.

To summarize this section on the classroom procedures for administering the CAT, we have seen a wide variety of teacher actions within a consistent sequence of major activities. The seating, materials, reading of directions, and timed response periods remain invariable, as would be expected in this situation. The sources of variation in the process are in the pre-instructions given by teachers, particularly in test taking strategies, in the extent to which they explain the printed test directions, the extent to which they provide encouragement and

praise, and particularly in their proctoring activities. Finally, at least some of this variation appears to be linked to teachers' opinions about the test itself.

CONCLUSIONS

In conclusion, we have found considerable variation in teachers' opinions about standardized achievement tests and in their classroom procedures for administering them. Some teachers use test results to get a general picture of their students or to group them, others find no use for the tests at all. Half of the teachers studied believed the CAT to be "cruel." Such opinions about the test were reflected in the teachers' classroom procedures for administering it: these procedures showed variation among teachers in instructions, explanations, and in the frequency of student/teacher interactions during the testing.

The variation in these areas, however, is overshadowed by the relatively consistent role which achievement tests play in the teacher's overall strategy for evaluating student progress - almost none. The relative insignificance of achievement tests for teachers' assessment of students is due in part to the teachers' belief that such tests are not particularly valid evaluation techniques. Teachers perceive the standardized test to represent an unusually formal (and therefore potentially

suspect) testing situation. More importantly, they perceive the standardized test as being unrelated to the curriculum, and thus of little use for them.

It might be (though this is by no means certain) that teachers would make more use of standardized achievement tests if these tests more closely matched their program of classroom instruction. Providing this match is no simple task, however. The very nature of a normed, standardized test prohibits it from being individually tailored for classes or even school districts. Conversely, different districts, schools and classrooms have different needs, and changing curricula to "teach to the test" is not likely to be a satisfactory solution. Teachers will need to see some convergence between curriculum and test, however, before they will put much faith in test results.

Table 1. Teacher Rankings of Test Validity
(1 = most valid 8 = least valid)

Tests	Teachers					
	A	B	C	D	E	F
Observation	8	4	4	1	1	1
Discussion	5	5	3	2	3	2
Classwork	4	3	2	3	5	3
Teacher-made Tests	1	1	1	4	4	4
Program Tests	2	2	5	5	2	5
Competency Tests						
a. Interim	6	6	8	6	6	6
b. Summative Tests	7	7	6	7	7	7
California Achievement Test	3	8	7	8	8	8

Table 2. Responses to Opinion Statements

	disagree strongly (1)	disagree (2)	neutral (3)	agree (4)	agree strongly (5)	AVG
The CAT is mass cruelty	1	2		1	2	3.17
There is no better way to assess students than by talking to them		2	1		3	3.67
Tests provide an objective measurement of students' knowledge			2	3	1	3.83

Figure 1. Taxonomy of Assessment Techniques

Observation		Watching	-in class -on playground -among selves
		Interacting	-ask questions -oral tests -guidance with work
Assigned Work	Classwork	Individual	-dittos -workbooks -program pages -directed exercises -small group activities
		Group(s)	-discussion -oral reading -board work -response cards -games
	Homework		-dittos -workbook -directed exercises
	Projects/Reports		-written -oral -graphic
Tests	Teacher-made		-formal tests -informal quizzes
	Program		-pretest/posttest -unit -reassessment
	Diagnostic		-Spache -Houghton-Mifflin (Reading) -corrective reading
	District CUES		-Summative -Interim
	Standardized		-California Achievement Test

Figure 2. Multidimensional Scaling of Evaluation Techniques

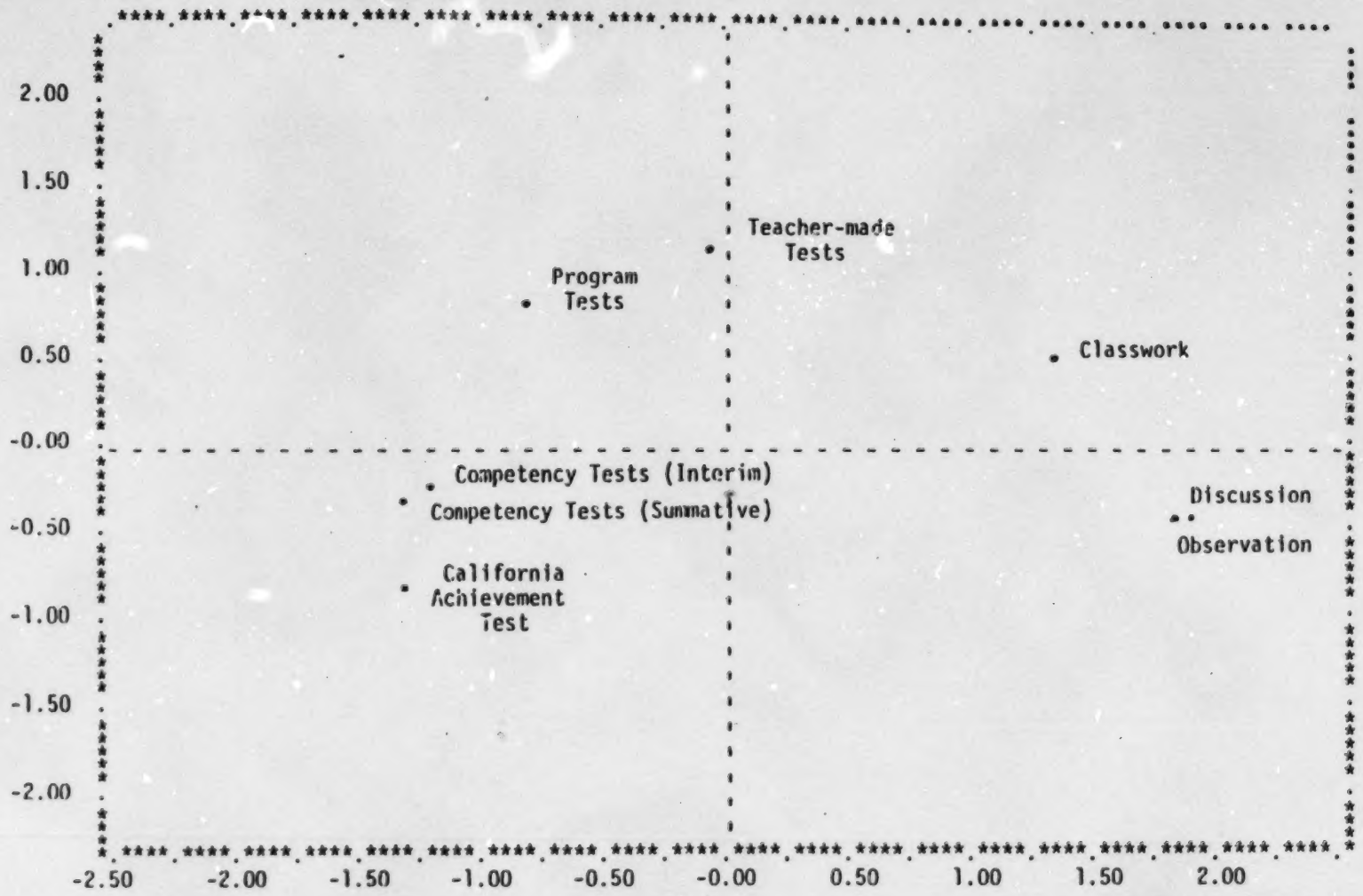
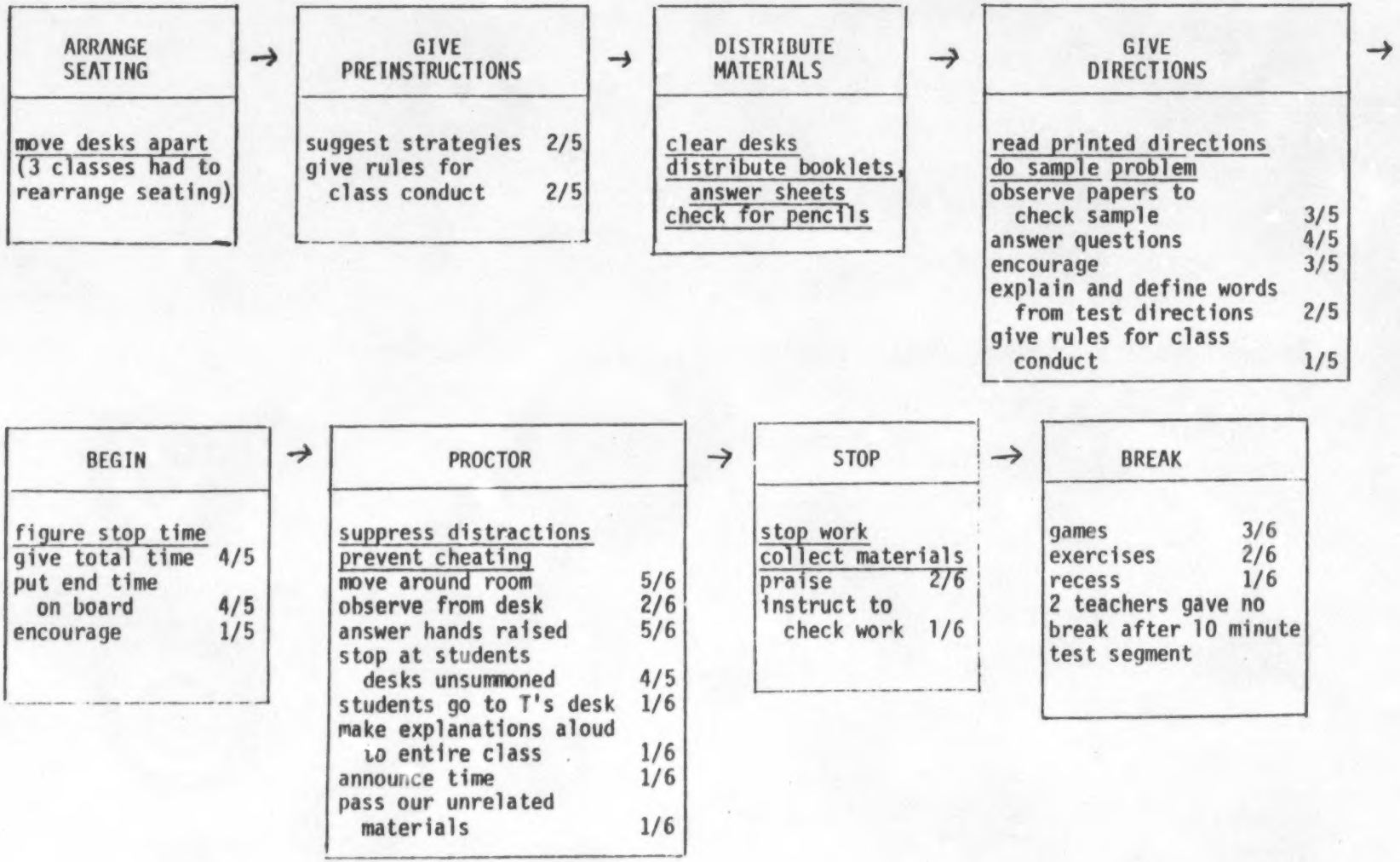


Figure 3. Classroom Procedure for Administering the California Achievement Test



References

Cole, N. S. & Nitko, A. J. Instrumentation and bias: Issues in selecting measures for educational evaluation. Paper presented at the National Symposium on Educational Research, Johns Hopkins University, November 1979.

Hanson, R. A., and Schutz, R. E. A new look at schooling effects from programmatic research and development. In D. Mann (Ed.), Making change happen. New York: Teacher College Press, 1978.

Leinhardt, G., & Seewals, A. M. Overlap: What's tested, what's taught? Unpublished manuscript, November, 1980. (Available from Learning Research and Development Center, University of Pittsburgh)

Rosenshine, B. Academic engaged minutes, content covered, and direct instruction. Unpublished manuscript, University of Illinois at Urbana-Champaign, 1978.

Salmon-Cox, L. Teachers and tests: What's really happening? Paper presented at the annual meeting of the American Educational Research Association, Boston, April 1980.

Spradley, J. P. The ethnographic interview. New York: Holt, Rinehart & Winston, 1979.

Yeh, J. Test use in the schools. Los Angeles, California: Center for the Study of Evaluation, University of California, Los Angeles, 1978.